PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

A14

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G01N 33/543, C07K 2/00, 14/00, C12Q 1/68, C12N 15/09, 15/70, 15/74 — A1

(11) International Publication Number: WO 96/41180

(43) International Publication Date: 19 December 1996 (19.12.96)

(21) International Application Number: PCT/US96/09383

(22) International Filing Date: 6 June 1996 (06.06.96)

(30) Priority Data:
08/479,660 7 June 1995 (07.06.95) US

(71) Applicant: PHARMACEUTICAL PEPTIDES, INC. [US/US]; 1 Hampshire Street, Cambridge, MA 02139-1572 (US).

(72) Inventors: BENJAMIN, Howard; 410 Marrett Road, Lexington, MA 02173 (US). SIGNER, Ethan; Apartment 2, 20 Forest Street, Cambridge, MA 02140 (US). GEFTER, Malcolm; 46 Baker Bridge Road, Lincoln, MA 01773 (US).

(74) Agent: GREER, Helen; Wolf, Greenfield & Sacks P.C., 600 Atlantic Avenue, Boston, MA 02210 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: ANCHOR LIBRARIES AND IDENTIFICATION OF PEPTIDE BINDING SEQUENCES

(57) Abstract

An anchor library is described. A collection of recombinant vectors having a nucleic acid encoding a displayed peptide sequence is provided. The displayed peptide sequence of each of the vectors comprises $X^1(Y^1)c^1X^2(Y^2)c^2X^3(Y^3)c^3X^4$, wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein X^1 and X^4 are each attached to an amino acid residue that flanks the displayed peptide sequence. Preferably, at least about 10^5 to about 10^8 permutations of all possible permutations of the displayed peptide sequence are present in the anchor library. Preferably, the library does not contain more than about 10 % of displayed peptide sequences different from the first mentioned displayed peptide sequences. Also described are methods of making anchor libraries and methods of using anchor libraries to identify a peptide sequence that binds to a target. Recombinant vectors, filamentous phage, nucleic acid molecules and proteins are also provided.

ANCHOR LIBRARIES AND IDENTIFICATION OF PEPTIDE BINDING SEQUENCES

Field of the Invention

This invention relates to anchor libraries and to methods of using anchor libraries to identify peptide sequences that bind to a target molecule.

Background of the Invention

The identification of peptides which bind to target molecules which are involved in various physiological functions, can have significant implications for the diagnosis and/or treatment of various abnormal or diseased conditions. For example, a binding peptide might modulate the original activity of the target molecule and therefore be useful as a drug.

The use of standard libraries to identify peptide sequences which specifically bind to target molecules is generally limited to pre-existing natural sequences from the organism which is the source of the DNA. More recently, libraries have been described which have clones containing short synthetic random coding sequences. See, e.g., Scott and Smith, Science 249:386-390 (1990); Cwirla et al., Proc. Natl. Acad. Sci. USA 87:6378-6382 (1990); Devlin et al., Science 249:404-406 (1990). These libraries are mixtures of filamentous phage clones, each displaying a random peptide sequence on the virion surface. In these types of libraries, the random amino acids are contiguous. The size of the peptides that can be screened for binding peptides in such contiguous random amino acid libraries is limited, in that as the size of the peptides increases, at some point it is not feasible to adequately search such a library since there are too many clones required to cover all possible permutations of the random amino acids in the peptides.

Summary of the Invention

It is an object of the invention to identify peptide sequences that bind to specific target molecules.

It is another object of the invention to identify amino acid residues in a peptide that are important contacts between the peptide and a target molecule.

It is another object of the invention to determine where amino acid residues in a peptide that are important contacts between the peptide and a target molecule, are best positioned within the peptide.

It is another object of the invention to use an anchor library in which the random amino acid residues of the library are not continuous, for identifying amino acid residues in a peptide that are important contacts between the peptide and a target molecule.

It is another object of the invention to use an anchor library in which the random amino acid residues of the library are distributed throughout a much larger peptide domain consisting of random glycine and/or alanine residues, for identifying amino acid residues in a peptide that are important contacts between the peptide and a target molecule.

It is another object of the invention to search large peptide phage display libraries of, e.g., 16 mers, for a reduced number of essential amino acid residue contacts, e.g., four, between the peptide and a target molecule.

It is another object of the invention to identify a consensus sequence of a defined number of amino acid residues in any configuration of spacer amino acids, that are important contacts between a peptide and a target molecule.

It is yet another object of the invention to use a known core binding sequence on a peptide which binds to a target molecule, and identify surrounding amino acid residues which are additional important contacts between the peptide and the target molecule.

Still another object of the invention is to identify cysteine residues on a peptide which can form disulfide bridges and thereby increase the binding affinity of the peptide with a target molecule.

According to the invention, an anchor library is provided. The anchor library comprises a collection of recombinant vectors. e.g., viruses, phage, e.g., filamentous phage, plasmids or cosmids. Each of the vectors has a nucleic acid sequence inserted in a gene, e.g., a coat protein gene, e.g., gene III or gene VIII, thioredoxin, staphnuclease, lac repressor, gal4 or an antibody. The nucleic acid sequence encodes a displayed peptide sequence, e.g., displayed on the surface of a virion, cell, spore or gene product, which comprises:

$$X^1(Y^1)_{C^1}X^2(Y^2)_{C^2}X^3(Y^3)_{C^3}X^4$$

wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively, c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of

c^1, c^2 and c^3 preferably is 0 to about 20, more preferably is 0 to about 10, even more preferably is 0 to about 6, or most preferably is 0 to about 4, wherein X^1 and X^4 are each attached to an amino acid residue that flanks the displayed peptide sequence. In certain embodiments, at least about 10^5 to about 10^8 permutations of all possible permutations of the displayed peptide sequence are present in the anchor library. In other embodiments, the library does not contain more than about 10%, or more than about 1%, or more than about 0.1%, of displayed peptide sequences different from the first mentioned displayed peptide sequences.

Another aspect of the invention is where each Y^1, Y^2 and Y^3 is any specified amino acid or combination of specified amino acids, e.g., alanine or cysteine or a combination of alanine and cysteine; or glycine or cysteine or a combination of glycine and cysteine.

In certain embodiments, the displayed peptide sequence further has at least one core binding sequence which is preferably about 1 to about 20 amino acid residues in length, more preferably about 4 to about 10, and most preferably is 6. The core binding sequence can be in addition to, or a replacement for, other amino acids in the displayed peptide sequence. Variations include the presence of more than one core binding sequence in the displayed peptide sequence, where, e.g., the core binding sequences can be adjacent, or not adjacent, to each other, and where they can be, e.g., identical or not identical to each other.

In other embodiments, the displayed peptide sequence further has at least one constraint, e.g., a crosslink, e.g., a disulfide bond, e.g., from the presence of a cysteine residue; a stacking interaction; a positive or negative charge; hydrophobicity; hydrophilicity; a structural motif, e.g., a zinc finger formation, a leucine zipper, or a β-turn structure, e.g., from the presence of the sequence asp gly or pro gly; or combinations thereof. Cysteine residues can be in addition to, or a replacement for, other amino acids in the displayed peptide sequence.

Another aspect of the invention is a method of making an anchor library. A collection of nucleic acid sequences is synthesized. The nucleic acid sequences are inserted into vectors to give recombinant vectors and the recombinant vectors are introduced into a host. The host having the recombinant vectors is propagated so as to result in a collection of recombinant vectors, each of which has a nucleic acid sequence from the collection of nucleic acid sequences which encodes a displayed peptide sequence comprising:

$$X^1(Y^1)_{c^1}X^2(Y^2)_{c^2}X^3(Y^3)_{c^3}X^4.$$

Another aspect of the invention is a method of using an anchor library to identify a peptide sequence that binds to a target. An anchor library having a collection of recombinant vectors is provided. Each of the recombinant vectors has a nucleic acid sequence which encodes a displayed peptide sequence comprising:

$$X^1(Y^1)_{C^1}X^2(Y^2)_{C^2}X^3(Y^3)_{C^3}X^4.$$

Expression and display of the peptide sequence is permitted. The anchor library is contacted with the target, e.g., proteinaceous or non-proteinaceous molecules, e.g., ligands, receptors, hormones, cytokines, antibodies, antigens, enzymes, enzyme substrates or viruses, under conditions in which the displayed peptide sequence binds to the target, and the displayed peptide sequence which binds to the target is identified, e.g., by sequencing the nucleic acid sequence on the recombinant vector which encodes for the displayed peptide sequence. Preferably, the identified displayed peptide sequence is synthesized.

The invention also provides for a peptide which is identified by use of an anchor library, in which the peptide is useful as a diagnostic or therapeutic product in that the peptide is able to bind to a target molecule which is involved in a physiological process.

Other aspects of the invention include, e.g., a collection of recombinant DNA molecules encoding peptide sequences having a plurality of different binding domains; a recombinant filamentous phage having a displayed peptide sequence with known binding properties and which is foreign to the filamentous phage; a recombinant vector having a nucleic acid sequence inserted in a gene, the nucleic acid sequence encoding a displayed peptide sequence having known binding properties; a recombinant nucleic acid molecule having a nucleic acid sequence inserted in a gene, the nucleic acid sequence encoding a displayed peptide sequence having known binding properties; and a recombinant protein having a displayed peptide sequence having known binding properties.

The above and other objects. features and advantages of the present invention will be better understood from the following specification.

Detailed Description

This invention provides an anchor library. The anchor library comprises a collection of recombinant vectors, each of which has a nucleic acid sequence inserted in a gene. The nucleic acid sequence encodes a displayed peptide sequence which comprises:

$$X^1(Y^1)_{c^1}X^2(Y^2)_{c^2}X^3(Y^3)_{c^3}X^4$$

wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively, c^1, c^2 and c^3 amino acids residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein X^1 and X^4 are each attached to an amino acid residue that flanks the displayed peptide sequence. In certain embodiments at least about 10^5 to about 10^8 permutations of all possible permutations of the displayed peptide sequence are present in the anchor library. In other embodiments, the library does not contain more than about 10%, or more than about 1%, or more than about 0.1% of displayed peptide sequences different from the first mentioned displayed peptide sequences.

By anchor library is meant a library in which the recombinant vectors have nucleic acid sequences which code for peptide sequences with random amino acids in which the random amino acids are not continuous. An anchor library is thus distinguishable from other random amino acid libraries in which all random amino acids in the peptide sequence of interest are contiguous. In anchor libraries, a given number of random amino acids are distributed throughout a larger peptide domain consisting of specifically designated amino acid residues. Anchor libraries are meant to include, e.g., external libraries, e.g., phage display libraries, and internal libraries, e.g., plasmid libraries. Chemical libraries can be anchor libraries.

Vectors are meant to include, e.g., phage, viruses, plasmids, cosmids, or any other suitable vector known to those skilled in the art. The vector has a gene, native or foreign, which is able to tolerate insertion of a foreign peptide into the gene product of the gene. By gene is meant an intact gene or fragment thereof. In the invention, the expressed gene product contains the inserted peptide.

For certain embodiments of this invention, e.g., where phage display libraries are employed, the preferred vectors are filamentous phage, though other vectors can be used.

Filamentous phage are single stranded DNA phage having coat proteins. Preferably, the gene that the nucleic acid sequence is inserted into is a coat protein gene of the filamentous phage. Preferred coat proteins are gene III or gene VIII coat proteins. Insertion of a foreign peptide into a coat protein gene results in the display of the foreign peptide on the surface of the phage. Insertion into any other gene product in which the inserted peptide is displayed can also be used in this invention. Examples of filamentous phage vectors which can be used in this invention are fUSE vectors, e.g., fUSE1, fUSE2, fUSE3 and fUSE5, in which the insertion is just downstream of the pIII signal peptide. Smith and Scott, Methods in Enzymology 217:228-257 (1993).

In other embodiments, e.g., where internal libraries are employed, the preferred vectors are plasmids, though other vectors can be used. The gene that the nucleic acid is inserted into is a gene which also results in display of the inserted peptide sequence. The gene can encode for an exported or non-exported gene product. Preferred genes include, e.g., thioredoxin, staphnuclease, lac repressor, gal4 or an antibody.

By recombinant vector is meant a vector having a nucleic acid sequence which is not normally present in the vector. The nucleic acid sequence is inserted into a gene present on the vector. Insertion of a nucleic acid into a gene is meant to include insertion within the gene or immediately 5′ or 3′ to, respectively, the beginning or end of the gene, such that when expressed, a fusion gene product is made. The nucleic acid sequence that is inserted includes, e.g., a synthesized nucleic acid sequence or a fragment of another nucleic acid molecule. The nucleic acid sequence encodes a displayed peptide sequence.

By displayed peptide sequence is meant a peptide sequence that is on the surface of, e.g., a virion, e.g. a phage or virus, a cell, a spore, or an expressed gene product. It is preferable to have the displayed peptide displayed such that it is able to bind to added target molecules. A displayed peptide sequence can be identical to, or not identical to, a naturally occurring peptide sequence.

The displayed peptide sequence can vary in size. As the size increases, the complexity of the anchor library increases, such that at some point a complete library is not obtainable. Complete libraries or incomplete libraries can be used in this invention. In certain embodiments, the complexity of the anchor library is at least about 10^8 to about 10^{11}. Preferably, the complexity is at least about 10^9. It is preferred that the total size of the displayed peptide sequence (the random amino acids plus the spacer amino acids) should not be greater than about 100 amino acids long, more preferably not greater than about 50 amino acids long, and most preferably not

greater than about 25 amino acids long. A particularly preferred library is made up of displayed peptides in which the longest of the peptides is 16 amino acids, i.e., a 16 mer library.

In large standard libraries, e.g., of 16 mers or greater, it is ordinarily not possible to search a library which contains all possible combinations of the 16 random amino acids. A major advantage of the anchor libraries of this invention is that these large libraries can be searched by looking for a reduced number of essential amino acid contacts between the peptides and the target. Preferably, the number of essential amino acid contacts should be sufficient to achieve micromolar binding. Preferably, the reduced number of essential contacts is about three to about ten, and most preferably it is about four. See Example 4. Thus, e.g., the number of combinations of four amino acid residue contacts in a 16 mer library is much less than the total number of combinations of all 16 amino acids in a 16 mer library, and therefore, this invention makes it possible to determine four important contact amino acids in a peptide of 16 amino acids in length, as opposed to standard screening of standard libraries in which such determinations cannot ordinarily be made.

In one embodiment of the invention, the displayed peptide sequence comprises

$$X^1(Y^1)_{C^1}X^2(Y^2)_{C^2}X^3(Y^3)_{C^3}X^4.$$

X^1, X^2, X^3 and X^4 are amino acid residues, each of which can be the same or different from any one of the others. Preferably, the amino acids are chosen from the 20 amino acids commonly found in naturally occurring proteins.

Y^1, Y^2 and Y^3 can be any specified amino acid residue or combination of specified amino acid residues, and each of the Ys, if present, can be the same or different from any one of the others. Preferably, the amino acids are spacer amino acids which will not significantly interfere with the binding between the peptide sequence and a target molecule. It is preferable to use combinations of two or more amino acids for the Y amino acids in a given library so as to reduce any limitations in the conformations of the displayed peptide that might be imposed by use of only one given amino acid. Most preferably, glycine and alanine residues are used in combination in the library. Glycine and alanine are small side chain amino acids that appear to act more as blanks than interfering contacts. In other embodiments, the Y amino acids can be amino acids which are chosen because they do significantly affect in some way the binding

between the peptide sequence and a target molecule. For example, glycine and cysteine residues can be used in combination, or alanine and cysteine residues can be used in combination.

Y^1 , Y^2 and Y^3 are, respectively c^1, c^2 and c^3 amino acid residues long. c^1, c^2 and c^3 can be the same or different from any one of the others. Preferably, each of c^1, c^2 and c^3 is 0 to about 20, more preferably is 0 to about 10, even more preferably is 0 to about 6, and most preferably is 0 to about 4.

For example, in an anchor library where each of the c's are 0 to 4, and the Y's are a combination of glycine and alanine, the minimal structure of the peptide sequence is 4 amino acids long (where each of c^1, c^2 and c^3 is 0):

$$X^1X^2X^3X^4,$$

and the maximal structure of the peptide sequence is 16 amino acids long (where each of c^1, c^2 and c^3 is 4):

$$X^1(G/A)(G/A)(G/A)(G/A)X^2(G/A)(G/A)(G/A)(G/A)X^3(G/A)(G/A)(G/A)(G/A)X^4,$$

where (G/A) is a glycine or alanine residue. This anchor library also contains all other in-between permutations of c, e.g., where c^1 is 0, c^2 is 1 and c^3 is 1; where c^1 is 1, c^2 is 1 and c^3 is 1; where c^1 is 2, c^2 is 1 and c^3 is 1; etc. All possible permutations of alanine and glycine for each of the designated c values are also included in this anchor library.

It is preferred that all possible permutations of the displayed sequence are present, that is, all combinations of c values and all combinations of, e.g., alanine and/or glycine, for each of the c values. In other embodiments, at least about 10^5 to about 10^8 permutations of all possible permutations are present in the anchor library, or at least about 10^4 permutations of all possible permutations are present in the anchor library, or at least about 10^5 permutations of all possible permutations are present in the anchor library, or at least about 10^6 permutations of all possible permutations are present in the anchor library, or at least about 10^7 permutations of all possible permutations are present in the anchor library, or at least about 10^8 permutations of all possible permutations are present in the anchor library, or at least about 10^9 permutations of all possible permutations are present in the anchor library.

In certain embodiments, the library does not contain more than about 10% of displayed peptide sequences different from the first mentioned displayed peptide sequences. In other embodiments, the library does not contain more than about 1% of displayed peptide sequences different from the first mentioned displayed peptide sequences. And in yet other embodiments, the library does not contain more than about 0.1% of displayed peptide sequences different from the first mentioned displayed peptide sequences.

In certain embodiments of the invention, the displayed peptide can have additional units of $X(Y)_c$. For example, it can have preferably about 1 to about 10 additional units, more preferably about 1 to about 5 additional units, and most preferably about 1 to about 3 additional units. In other embodiments, one or more additional units of X alone or $(Y)_c$ alone can be present.

In yet other embodiments of the invention, the anchor libraries described above can have at least one core binding sequence, denoted by B, of p amino acid residues in length. B can be any size, e.g., from a single amino acid to the size of a gene. Preferably, p is about 1 to about 20, more preferably p is about 4 to about 10, and most preferably p is about 6. By core binding sequence is meant a peptide sequence which is known to bind to a target molecule. In certain embodiments, the core binding sequence is additional to the amino acid residues of the displayed peptide sequences described above. In such libraries, the core binding sequence can be positioned on the NH_2-terminal or COOH-terminal side of any of the X^1, X^2, X^3 or X^4 amino acid residues, or on the NH_2-terminal or COOH-terminal side of any of the Y, e.g., alanine or glycine, residues. In other embodiments, at least one of the X residues is replaced with the core binding sequence. In yet other embodiments, at least one of the Y residues, e.g., one of the alanine or glycine residues, is replaced with a core binding sequence. Inclusion of a known core binding sequence in the anchor library allows identification of surrounding amino acid residues which are additional important contacts between the peptide and the target molecule. The invention thus allows identification of better binding sequences by identifying additional amino acids surrounding the core binding sequence which in combination with the known core binding sequence exhibit enhanced binding as compared to the known core binding sequence alone.

In certain embodiments, more than one known binding sequence is present in each of the displayed peptide sequences of the anchor library. These multiple known binding sequences can be adjacent to, or not adjacent to, each other, and can be identical to, or not identical to, each other.

In certain embodiments, the anchor libraries have at least one constraint imposed upon the displayed peptide sequence. A constraint includes, e.g., a crosslink, a stacking interaction, a positive or negative charge, hydrophobicity, hydrophilicity, a structural motif and combinations thereof. In certain embodiments, more than one constraint is present in each of the displayed peptide sequences of the anchor library. These multiple constraints can be adjacent to, or not adjacent to, each other, and can be identical to, or not identical to, each other.

A crosslink includes, e.g., a disulfide bond. In certain embodiments, the displayed peptide has at least one cysteine residue. The cysteine residue can be, e.g., additional to the amino acid residues of the displayed peptide sequences described above. In such libraries, the cysteine residue can be positioned on the NH_2-terminal or COOH-terminal side of any of the X^1, X^2, X^3 or X^4 amino acid residues, or on the NH_2-terminal or COOH-terminal side of any of the Y, e.g., alanine or glycine, residues. In other embodiments, at least one of the X residues is a cysteine residue. In yet other embodiments, at least one of the Y residues, e.g., one of the alanine or glycine residues, is replaced with a cysteine residue. Multiple cysteines can be present in each of the peptides so as to form potential disulfide bonds within a random series. Disulfide bonds can be formed within the displayed peptide sequence itself or between the displayed peptide sequence and the target molecule.

A structural motif includes, e.g., a zinc finger formation, a leucine zipper, and a β-turn structure in the peptide. The sequences asp gly or pro gly are likely to induce β-turns, either alone or in combination with, e.g., a disulfide bond.

In other embodiments, the anchor libraries can be constructed to have both a core binding sequence and a constraint, e.g., at least one cysteine residue. In one such embodiment, at least one of the X residues can be, e.g., either a cysteine or a glycine such that the displayed peptide sequence is:

$$(C/G)(Y^1)_{C1}(C/G)(Y^2)_{C2}B(C/G)(Y^3)_{C3}(C/G)$$

where (C/G) is a cysteine or glycine residue. In such a library, multiple cysteines are present so as to form potential disulfide bonds within a random series.

In yet other embodiments, the displayed peptide sequence comprises:

$$X^1(Y^1)_{C1}X^2(Y^2)_{C2}X^3(Y^3)_{C3}X^4$$

wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a core binding sequence B of p amino acid residues in length or a combination of alanine and glycine or alanine and B or glycine and B.

And in yet other embodiments, the displayed peptide sequence comprises:

$$Z^1(Y^1)_{C^1}Z^2(Y^2)_{C^2}Z^3(Y^3)_{C^3}Z^4$$

wherein each Z^1, Z^2, Z^3 and Z^4 is an amino acid residue or a core binding sequence B of p amino acid residues in length and any of Z^1, Z^2, Z^3 and Z^4 can be the same or different from any one other, and wherein Z^1 and Z^4 are each attached to an amino acid residue that flanks the displayed peptide sequence.

Other embodiments include anchor libraries constructed with other configurations of combinations between X residues and/or Y residues and/or B sequences and/or cysteine residues and/or other constraints, as is obvious to those skilled in the art.

The invention also includes a method of making the anchor libraries described above. A collection of nucleic acid sequences is synthesized and inserted into vectors to give recombinant vectors. These recombinant vectors are introduced into a host. The host having the recombinant vectors is propagated so as to result in a collection of recombinant vectors, each of the recombinant vectors having a nucleic acid sequence from the collection of nucleic acid sequences which encodes a displayed peptide sequence. The peptide sequence is any of the peptide sequences discussed above, e.g., $X^1(Y^1)_{C^1}X^2(Y^2)_{C^2}X^3(Y^3)_{C^3}X^4$, with or without at least one core binding sequence, and with or without at least one constraint, e.g., a cysteine residue. In certain embodiments, at least about 10^5 to about 10^8 permutations, or about 10^4 permutations, or about 10^5 permutations, or about 10^6 permutations, or about 10^7 permutations, or about 10^8 permutations, or about 10^9 permutations, of all possible permutations of the displayed peptide sequence are present in the anchor library. In other embodiments, the library does not contain more than about 10%, or more than about 1%, or more than about 0.1%, of displayed peptide sequences different from the first mentioned displayed peptide sequences.

The nucleic acids that encode the anchor library can be obtained by any method which produces the requisite permuted nucleic acids. For example, a split synthesis procedure can be used. See, e.g., Cormack and Struhl, Science 262:244-248 (1993). Examples 1 and 3 describe examples of using split synthesis to make nucleic acid inserts for anchor libraries.

The invention further includes a method of using the anchor libraries described above to identify a peptide sequence that binds to a target. An anchor library having a collection of recombinant vectors, each of which has a nucleic acid sequence which encodes a displayed peptide sequence, is provided. The displayed peptide sequence can be any of the peptide sequences discussed above, e.g., $X^1(Y^1)_{C^1}X^2(Y^2)_{C^2}X^3(Y^3)_{C^3}X^4$, with or without at least one core binding sequence, and with or without at least one constraint, e.g., a cysteine residue. Expression and display of the peptide sequence is permitted. The anchor library is contacted with the target under conditions in which the displayed peptide sequence binds to the target, and the displayed peptide sequence which binds to the target is identified.

Target is meant to include any molecule with which the displayed peptide sequence will bind. Targets include, e.g., proteinaceous and non-proteinaceous molecules. Examples of targets are ligands, receptors, hormones, cytokines, antibodies, antigens, enzymes, enzyme substrates and viruses. In some cases, the binding peptide modulates the original activity of the target molecule, and therefore can be useful as a drug. The target includes, e.g., drug antagonists and agonists. The binding peptides can be used, e.g., for diagnostic or therapeutic applications.

The contacting step can be done by any method in which the displayed peptide sequence will bind, directly or indirectly, to the target. These methods include, e.g., screens and selections. Preferably, an affinity purification method is used. Affinity purification includes, e.g., biopanning. For example, a phage anchor library having displayed peptide sequences is mixed with biotinylated target, resulting in phage:biotinylated target complex if a displayed peptide sequence binds to the target. The mixture is added to a streptavidin coated substance, e.g., beads or a petri plate. The resulting biotin-streptavidin bond allows isolation of the phage carrying peptide sequences that bind to the target., It is preferable to do multiple rounds of biopanning to reduce background. See Example 2.

Identification of the displayed peptide sequence includes, e.g., determining the sequence of amino acids that comprise the peptide. Identification can be accomplished, e.g., by amplifying the recombinant vector which has the nucleic acid sequence which encodes for the displayed peptide sequence which binds to the target, and sequencing the nucleic acid sequence by standard procedures known in the art to determine the displayed peptide sequence which binds to the target. If desired, the peptide thus identified can be synthesized using standard procedures known in the art and further tested for its ability to bind to the target in vitro and/or in cell-based, and/or animal models. See Example 2.

In a given anchor library, the ability to determine essential amino acid contacts between the displayed peptide and a target molecule is aided by the ability to observe conserved amino acid residues in the different displayed peptides which are able to bind to the target. Conserved amino acid residues are meant to include different DNA codons for the same amino acid or different DNA codons for functionally similar amino acids. The consensus is determined by comparing the sequence of the individual clones obtained from a library screen. It is preferable that the library have sufficient complexity in order to observe such a consensus.

Also included in the invention is a peptide identified by use of any of the anchor libraries described above in which the peptide is useful as a diagnostic or therapeutic product in that the peptide is able to bind to a target molecule which is involved in a physiological process. For example, the target molecule can be a receptor involved in inflammation, e.g., IL-1, or in prostate cancer, e.g., GnRH; or the target molecule can be an enzyme, e.g., a protease, e.g., HIV protease. By binding to these or other target molecules that are involved in various abnormal conditions or diseases, the binding peptides of this invention modulate the original activity of the target molecule and are therefore useful as diagnostic or therapeutic products.

The invention also includes a library which has a collection of nucleic acid molecules encoding peptides having random amino acids, the improvement comprising a library in which the random amino acids are not continuous so that the amino acids in the peptide that are important contacts for interaction between the peptide and a target molecule can be identified.

The invention also includes a library having a collection of nucleic acid molecules encoding peptides having random amino acids, the improvement comprising nucleic acid molecules encoding alanine or glycine or a combination of alanine and glycine residues in varying numbers acting as spacers between the random amino acids so that amino acid residues in a peptide that are important contacts for interaction between the peptide and a target molecule can be identified.

The invention further provides a collection of recombinant DNA molecules encoding peptide sequences having a plurality of different binding domains. The peptide sequences comprise: $X^1(Y^1)_{C^1}X^2(Y^2)_{C^2}X^3(Y^3)_{C^3}X^4$, wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein X^1 and X^4 are each

attached to an amino acid residue that flanks the peptide sequence, and wherein at least about 10^5 to about 10^8 permutations, or about 10^4 permutations, or about 10^5 permutations, or about 10^6 permutations, or about 10^7 permutations, or about 10^8 permutations, or about 10^9 permutations, of all possible permutations of the peptide sequence are present in the collection. In other embodiments, the collection does not contain more than about 10%, or more than about 1%, or more than about 0.1%, of displayed peptide sequences different from the first mentioned displayed peptide sequences. In certain embodiments, the peptide sequences are displayed on the surface of a biological material, e.g., a virus, phage, cell, spore or gene product.

The invention also includes a recombinant filamentous phage having a displayed peptide sequence with known binding properties. The displayed peptide sequence is foreign to the filamentous phage. The displayed peptide sequence comprises: $X^1(Y^1)_{C^1}X^2(Y^2)_{C^2}X^3(Y^3)_{C^3}X^4$, wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein X^1 and X^4 are each attached to an amino acid residue that flanks the displayed peptide sequence, and wherein the displayed peptide sequence is able to bind to a target. In certain embodiments, at least one of Y^1, Y^2 and Y^3 is at least about 20 amino acid residues long, preferably is at least about 10 amino acid residues long, more preferably is at least about 6 amino acid residues long, even more preferably is at least about 4 amino acid residues long, more preferably yet is at least about 3 amino acid residues long, more preferably yet is at least about 2 amino acid residues long, and most preferably is at least about 1 amino acid residue long.

The invention also includes a recombinant vector having a nucleic acid sequence inserted in a gene. The nucleic acid sequence encodes a displayed peptide sequence having known binding properties. The displayed peptide sequence comprises: $X^1(Y^1)_{C^1}X^2(Y^2)_{C^2}X^3(Y^3)_{C^3}X^4$, wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein X^1 and X^4 are each attached to an amino acid residue that flanks the displayed peptide sequence, and wherein the displayed peptide sequence is able to

bind to a target. In certain embodiments, at least one of Y^1, Y^2 and Y^3 is at least about 20 amino acid residues long, preferably is at least about 10 amino acid residues long, more preferably is at least about 6 amino acid residues long, even more preferably is at least about 4 amino acid residues long, more preferably yet is at least about 3 amino acid residues long, more preferably yet is at least about 2 amino acid residues long, and most preferably is at least about 1 amino acid residue long.

The invention also includes a recombinant nucleic acid molecule having a nucleic acid sequence inserted in a gene. The nucleic acid sequence encodes a displayed peptide sequence having known binding properties. The displayed peptide sequence comprises: $X^1(Y^1)_{C^1}X^2(Y^2)_{C^2}X^3(Y^3)_{C^3}X^4$, wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein X^1 and X^4 are each attached to an amino acid residue that flanks the displayed peptide sequence, and wherein the displayed peptide sequence is able to bind to a target. In certain embodiments, at least one of Y^1, Y^2 and Y^3 is at least about 20 amino acid residues long, preferably is at least about 10 amino acid residues long, more preferably is at least about 6 amino acid residues long, more preferably is at least about 4 amino acid residues long, more preferably yet is at least about 3 amino acid residues long, more preferably yet is at least about 2 amino acid residues long, and most preferably is at least about 1 amino acid residue long.

The invention further includes a recombinant protein having a displayed peptide sequence having known binding properties. The displayed peptide sequence comprises: $X^1(Y^1)_{C^1}X^2(Y^2)_{C^2}X^3(Y^3)_{C^3}X^4$, wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein X^1 and X^4 are each attached to an amino acid residue that flanks the displayed peptide sequence, and wherein the displayed peptide sequence is able to bind to a target. In certain embodiments, at least one of Y^1, Y^2 and Y^3 is at least about 20 amino acid residues long, preferably is at least about 10 amino acid residues long, more preferably is at least about 6 amino acid residues long, even more preferably is at least

about 4 amino acid residues long, more preferably yet is at least about 3 amino acid residues long, more preferably yet is at least about 2 amino acid residues long, and most preferably is at least about 1 amino acid residue long.

EXAMPLES

Example 1: Construction of a Phage Anchor Library

This example illustrates the construction of a phage anchor library having random amino acid codons distributed throughout a domain of alanine and/or glycine codons. Standard cloning techniques known to those skilled in the art were used.

(a) Vector Preparation

30 μg of Fuse5 (Smith and Scott, Methods in Enzymology 217:228-257 (1993)) was cleaved with 200 units of endonuclease Sfi I in 500 μl of NEB #2 restriction buffer for 10 hours. The reaction was terminated with addition of 15 mM EDTA, followed by phenol and chloroform extractions. The DNA was recovered by isopropanol precipitation, resuspended in 500 μl of TE, and recovered by EtOH precipitation.

(b) Insert Preparations

The anchor insert used in the library was synthesized as a single stranded oligomer using split synthesis. See, e.g., Cormack and Struhl, Science 262:244-248 (1993). This process creates combinations of sequences which differ from each other.

Using split synthesis, five templates were synthesized and mixed three times to produce the anchor library:

1) GGGCTGCCGGGNNKNNK (Seq. ID No. 1)	
2) GGCTGCCGGGNNKGSNNNK (Seq. ID No. 2)	
3) GGGCTGCCGGGNNKGSNGSNNNK (Seq. ID No. 3)	COMBINE AND SPLIT
4) GGGCTGCCGGGNNKGSNGSNGSNNNK (Seq. ID No. 4)	
5) GGGCTGCCGGGNNKGSNGSNGSNGSNNNK (Seq. ID No. 5)	
6) NNK	
7) GSNNNK	
8) GSNGSNNNK	COMBINE AND SPLIT
9) GSNGSNGSNNNK (Seq. ID No. 6)	
10) GSNGSNGSNGSNNNK (Seq. ID No. 7)	
11) NNKGGTGGTGCTGCTG (Seq. ID No. 8)	
12) GSNNNKGGTGGTGCTGCTG (Seq. ID No. 9)	
13) GSNGSNNNKGGTGGTGCTGCTG (Seq. ID No. 10)	COMBINE
14) GSNGSNGSNNNKGGTGGTGCTGCTG (Seq. ID No. 11)	
15) GSNGSNGSNGSNNNKGGTGGTGCTGCTG (Seq. ID No. 12)	

N = equal mix of G, A, T, C
S = equal mix of G, C
K = equal mix of G, T

DNA was chemically synthesized such that column 1 contained the DNA sequence GGGCTGCCGGG (Seq. ID No. 13), followed by DNA encoding a random amino acid, NNK, followed by DNA encoding a second random amino acid, NNK. Column 2 encoded the DNA sequence GGGCTGCCGGG (Seq. ID No. 13), followed by a random amino acid codon, NNK, followed by either a glycine or alanine codon, GSN, and then followed by a random amino acid codon, NNK. Columns 3, 4 and 5 encoded the DNA sequence GGGCTGCCGGG (Seq. ID No. 13), followed by a random amino acid codon, NNK, followed by, respectively, 2, 3 and 4 glycine and/or alanine codons, GSN, and then followed by a random amino acid codon, NNK.

After synthesis of columns 1-5, the resins from the five columns were mixed, resulting in a pool of oligomers which contained two random amino acids separated by 0 to 4 glycine and/or alanine residues. This entire mixture was then split into 5 new columns, denoted 6-10. Each of these columns was subjected to further DNA synthesis, resulting in, respectively, codons for 0, 1, 2, 3 and 4 glycine and/or alanine residues, GSN, followed by a random amino acid, NNK. Because the additions of columns 6-10 were conducted on a mixture of resins from columns 1-5, the mixture of columns 6-10 resulted in oligomers that all have three random amino acids, such that the neighboring random amino acids are separated by 0 to 4 glycine and/or alanine residues.

One additional round of split synthesis was undertaken in which the mixtures of columns 6-10 were extended with 0 to 4 glycine and/or alanine residues, GSN, and one more additional random amino acid, NNK, followed by the sequence GGTGGTGCTGCTG (Seq. ID No. 14). The final mixture of these columns resulted in a series of oligomers with four random amino acids such that the neighboring random amino acids are separated by 0 to 4 glycine and/or alanine residues.

Two additional oligomers, pins, CCCGGCAGCCCCGT (Seq. ID No. 15) and CAGCACCACC (Seq. ID No. 16), were synthesized which hybridize to the anchor oligomers so as to reconstruct double stranded DNA near the termini of the insert with three single strand nucleotide overhangs corresponding to Sfi I overhangs.

The insert and pin oligomers were kinased at 10 μg/30 μl kinase buffer from NEB with 1 mM ATP at 37°C for 30 minutes, followed by inactivation at 68°C for 5 minutes. The anchor oligomer was annealed to the pin oligomers in 500 mM NaCl, 50 mM Tris pH 7.5 at 68°C for 10 minutes and cooled to room temperature over 30 minutes. Each of the oligomers was at 5 μM during the annealing.

It is noted that similar results can be obtained with other 5' and 3' flanking sequences on the anchor inserts, and with other corresponding pin sequences altered appropriately, as can be

chosen by those skilled in the art. Moreover, other restriction sites can be used as appropriate for any given vector, as is known to those skilled in the art.

(c) Vector Ligation

30 µg of DNA vector was ligated to assembled insert at 5 µg/ml vector and three-fold excess assembled insert in NEB ligation buffer with 100 units of T4 DNA ligase at 10°C for 16 hours. DNA was purified from ligation buffer by phenol and chloroform extractions, followed by EtOH precipitation and resuspension in TE.

(d) DNA Transformation

DNA was transformed into MC1061 (Wertman et al., Gene 49:253-262 (1986)) electrocompetent cells using 0.5 µg of DNA per 100 µl of cells using 0.2 cm electroporator cells and a BioRad electroporator set at 25 µF, 2.5 KV and 200 ohms. Shocked cells were recovered in SOC media, grown out at 37°C for 20 minutes and inoculated into LB containing 20 µg/ml tetracycline.

(e) Library Phage Isolation

Phage released from transformed cells were isolated after growing for 16 hours. Phage were separated from cells by centrifugation at 4°C at 4.2K for 30 min. In a Beckman J6, followed by a second centrifugation of the supernatant at 4.2K for 30 min. Phage were precipitated with the addition of 150 ml at 16.7% PEG/3.3 M NaCl per liter of supernatant. Mixed solutions were incubated at 4°C for 16 hours. Precipitated phage was collected at 4.2K in a J6 followed by resuspension in 40 ml of TBS. Resuspended phage were precipitated again with the addition of 4.5 ml of PEG solution for 4 hours. Phage were collected at 5K in a Beckman JA20 at 4°C. Phage were suspended in 7 ml of TBS and brought to 1.3 mg/ml density by the addition of 1 gm of CsCl per 2.226 gm of aqueous solution. Phage were subjected to equilibrium centrifugation in a type 80 rotor at 45K rpm for 40 hours. Phage bands were isolated, diluted 20 fold with TBS and pelleted at 40 K in a Type 50 rotor. Pellets were resuspended in 0.7 ml of TBS and used as is for biopanning at approximately 3×10^{13} phage/ml.

Example 2: Biopanning to Select for Peptide Binding Sequences

This example illustrates biopanning of the phage library obtained from Example 1 to select for displayed peptide sequences that bind to biotinylated IL-1B. The phage act as affinity-selectable vectors in that the displayed peptide binds specifically to immobilized IL-1B if the library contains a displayed peptide that can so interact with IL-1B.

(a) Binding

Biotinylated IL-1 (b-IL-1) (Yew et al., JBC 264(30):17691-17697 (1989)) is incubated with 1 x 10^{11} phage in 20 µl of TBS for 20 minutes at 22°C. The phage:(b-IL-1) complex is isolated from free phage by addition of streptavadin coated paramagnetic beads for an additional 10 minutes. Magnetic beads are collected by attraction with a magnet and washed with TBS containing 0.5% Tween-20 for a total of 7 washes over 30 minutes. The remaining phage that are bound to the beads (by way of b-IL-1 binding to streptavadin) are recovered by elution with 100 µl of 100 mM glycine pH 2.2 for 10 minutes. Eluted phage are neutralized with 1 M Tris base.

(b) Amplification

Eluted phage are amplified by infection into log phase K91 E. coli (Lyons and Zinder, Virology 49:45-60 (1972); Smith and Scott, Methods in Enzymology 217:228-257 (1993)) at an moi of 0.0001. Approximately 10^5 phage are amplified by plating on 10 LB agar petri dishes containing 20 µg/ml tetracycline. The phage released from infected cells, approximately 10^{12} phage, are harvested by washing the LB agar plates with LB, and purified as above through the two PEG precipitations and resuspended at 10^{13} phage/ml.

Amplified phage are further subjected to two additional rounds of biopanning using the binding and amplification conditions described above.

(c) Sequencing Inserts

After three rounds of biopanning, individual phage are isolated and sequenced to reveal the DNA sequence that encodes for the displayed peptide in the selected phage. Sequencing is done according to manufacturer's protocol for Sequenase 2.0 (United States Biochemical, Cleveland, OH 44122).

(d) Peptide Synthesis

Peptides representing affinity purified phage are synthesized (Research Genetics, Huntsville, AL 35801) and tested for their ability to bind IL-1 and effect IL-1 binding to IL-1 receptor in cell based and animal models. Slack et al., Biotechniques 10:1132-1138 (1989).

Example 3: Construction of a Phage Anchor Library Having Codons For a Known Core Peptide Binding Sequence

This example illustrates construction of a phage anchor library which has codons for a known core peptide binding sequence which binds to a target molecule, surrounded by random amino acid codons distributed throughout a domain of random alanine and/or glycine codons.

Construction of this type of library is similar to that described in Example 1, except that the oligomer constructs not only have the random amino acid codons and glycine and/or alanine codons, but also have nucleic acid sequences which code for a known core peptide binding sequence, denoted as B:

1)	GGGCTGCCGGGNNKNNK (Seq ID No. 1)	
2)	GGGCTGCCGGGNNKGNNNK (Seq. ID No. 2)	
3)	GGGCTGCCGGGNNKGSNGSNNNK (Seq. ID No. 3)	COMBINE AND SPLIT
4)	GGGCTGCCGGGNNKGSNGSNGSNNNK (Seq. ID No. 4)	
5)	GGGCTGCCGGGNNKGSNGSNGSNGSNNNK (Seq. ID No. 5)	
6)	BNNK	
7)	BGSNNNK	
8)	BGSNGSNNNK	COMBINE AND SPLIT
9)	BGSNGSNGSNNNK (Seq. ID No. 6)	
10)	BGSNGSNGSNGSNNNK (Seq. ID No. 7)	
11)	NNKGGTGGTGCTGCTG (Seq. ID No. 8)	
12)	GSNNNKGGTGGTGCTGCTG (Seq. ID No. 9)	
13)	GSNGSNNNKGGTGGTGCTGCTG (Seq. ID No. 10)	COMBINE
14)	GSNGSNGSNNNKGGTGGTGCTGCTG (Seq. ID No. 11)	
15)	GSNGSNGSNGSNNNKGGTGGTGCTGCTG (Seq. ID No. 12)	

The anchor library can also be constructed such that sequence B is located, e.g., before or after any of the other NNK or GSN codons.

Other anchor libraries, containing additions or substitutions of nucleic acid sequences, can be constructed using similar methods. For example, codons for cysteine, or any other specified amino acid or sequence of amino acids, can be substituted for the nucleic acid sequence coding for the core binding sequence B in the above-described split synthesis. Anchor libraries containing two or more core binding sequences, cysteines, or any other specified amino acid or sequence of amino acids, also can be constructed using similar procedures as described, except that the multiple additions are synthesized as part of the oligomers at multiple positions, e.g., each can be located before or after any of the NNK or GSN codons, as can be chosen by one skilled in the art.

Example 4: Four Amino Acid Residues in a Peptide Is Sufficient For Binding to a Target

This example illustrates that four amino acid residues in a peptide are sufficient for micromolar binding between the peptide and its target.

A hexamer phage library was constructed essentially as described for the anchor libraries, except the oligonucleotide was:
GGGCTGCCGGGNNKNNKNNKNNKNNKGGTGGTGCTGCTG (Seq. ID No. 18). The library was screened against an antibody to hCG by biopanning as described in Example 2. The phage that bound to the antibody contained the consensus sequence XaaThrProTrpXaaGln (Seq. ID No. 17), where X was not absolutely specified. Peptides were synthesized which corresponded to the identified sequences and the flanking amino acids found in the phage. These peptides had an IC50 of 4.5 μM compared to 10 nM for hCG. IC50 is equal to the concentration of peptide necessary to prevent 50% of hCG-I^{125} from binding to the antibody. Therefore, four amino acid residues were sufficient to result in μM binding.

Those skilled in the art will be able to ascertain, using no more than routine experimentation, many equivalents of the specific embodiments of the invention described herein. These and all other equivalents are intended to be encompassed by the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: PHARMACEUTICAL PEPTIDES, INC.
(B) STREET: ONE HAMPSHIRE STREET
(C) CITY: CAMBRIDGE
(D) STATE: MASSACHUSETTS
(E) COUNTRY: UNITED STATES OF AMERICA
(F) ZIP: 02139-1572

(ii) TITLE OF INVENTION: ANCHOR LIBRARIES AND IDENTIFICATION OF PEPTIDE BINDING SEQUENCES

(iii) NUMBER OF SEQUENCES: 18

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
(B) STREET: 600 Atlantic Avenue
(C) CITY: Boston
(D) STATE: Massachusetts
(E) COUNTRY: USA
(F) ZIP: 02210

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/479,660
(B) FILING DATE: 07-JUN-1995

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Greer, Helen
(B) REGISTRATION NUMBER: 36,816
(C) REFERENCE/DOCKET NUMBER: P0567/7000WO

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617) 720-3500
(B) TELEFAX: (617) 720-2441

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GGGCTGCCGG GNNKNNK 17

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GGGCTGCCGG GNNKGSNNNK 20

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGGCTGCCGG GNNKGSNGSN NNK 23

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGGCTGCCGG GNNKGSNGSN GSNNNK 26

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GGGCTGCCGG GNNKGSNGSN GSNGSNNNK 29

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GSNGSNGSNN NK 12

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GSNGSNGSNG SNNNK 15

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

NNKGGTGGTG CTGCTG 16

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GSNNNKGGTG GTGCTGCTG 19

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GSNGSNNNKG GTGGTGCTGC TG 22

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GSNGSNGSNN NKGGTGGTGC TGCTG 25

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GSNGSNGSNG SNNNKGGTGG TGCTGCTG 28

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GGGCTGCCGG G 11

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GGTGGTGCTG CTG 13

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CCCGGCAGCC CCGT 14

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CAGCACCACC 10

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Xaa Thr Pro Trp Xaa Gln
1 5

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GGGCTGCCGG GNNKNNKNNK NNKNNKNNKG GTGGTGCTGC TG 42

CLAIMS

1. An anchor library, comprising:

a collection of recombinant vectors,

each of said recombinant vectors having a nucleic acid sequence inserted in a gene, said nucleic acid sequence encoding a displayed peptide sequence,

said displayed peptide sequence of each of said vectors comprising

$$X^1(Y^1)_{c^1}X^2(Y^2)_{c^2}X^3(Y^3)_{c^3}X^4$$

wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein X^1 and X^4 are each attached to an amino acid residue that flanks said displayed peptide sequence, and

wherein at least about 10^5 to about 10^8 permutations of all possible permutations of said displayed peptide sequence are present in said anchor library.

2. The library of claim 1 wherein said library does not contain more than about 10% of displayed peptide sequences different from said first mentioned displayed peptide sequences.

3. The library of claim 1 wherein said library does not contain more than about 1% of displayed peptide sequences different from said first mentioned displayed peptide sequences.

4. The library of claim 1 wherein said library does not contain more than about 0.1% of displayed peptide sequences different from said first mentioned displayed peptide sequences.

5. The library of claim 1 wherein at least about 10^5 permutations of all possible permutations of said displayed peptide sequence are present in said anchor library.

6. The library of claim 1 wherein at least about 10^6 permutations of all possible permutations of said displayed peptide sequence are present in said anchor library.

7. The library of claim 1 wherein at least about 10^7 permutations of all possible permutations of said displayed peptide sequence are present in said anchor library.

8. The library of claim 1 wherein at least about 10^8 permutations of all possible permutations of said displayed peptide sequence are present in said anchor library.

9. The library of claim 1 wherein said vector is selected from the group consisting of a virus, phage, plasmid and cosmid.

10. The library of claim 1 wherein said vector is a filamentous phage.

11. The library of claim 10 wherein said gene that said nucleic acid sequence is inserted in is a coat protein gene of said filamentous phage.

12. The library of claim 10 wherein said gene that said nucleic acid sequence is inserted in is a filamentous phage gene selected from the group consisting of gene III and gene VIII.

13. The library of claim 10 wherein said gene that said nucleic acid sequence is inserted in is selected from the group consisting of thioredoxin, staphnuclease, lac repressor, gal4 and an antibody.

14. The library of claim 1 wherein said displayed peptide sequence is displayed on the surface of a virion.

15. The library of claim 1 wherein said displayed peptide sequence is displayed on the surface of a cell.

16. The library of claim 1 wherein said displayed peptide sequence is displayed on the surface of an expressed gene product.

17. The library of claim 1 wherein each of said c^1, c^2 and c^3 is 0 to about 10.

18. The library of claim 1 wherein each of said c^1, c^2 and c^3 is 0 to about 6.

19. The library of claim 1 wherein each of said c^1, c^2 and c^3 is 0 to about 4.

20. The library of claim 1 further comprising about 1 to about 10 additional units of $X(Y)_C$.

21. The library of claim 1 wherein said displayed peptide sequence is not identical to a naturally occurring peptide sequence.

22. The library of claim 1 wherein said displayed peptide sequence is identical to a naturally occurring peptide sequence.

23. The library of claim 1 wherein said displayed peptide sequence further comprises at least one B, said B being a core binding sequence of p amino acid residues in length.

24. The library of claim 23 wherein p is about 1 to about 20.

25. The library of claim 23 wherein p is about 4 to about 10.

26. The library of claim 23 wherein p is about 6.

27. The library of claim 23 wherein said B is selected from the group consisting of said B being on the NH_2-terminal side of any of said X^1, X^2, X^3 or X^4 amino acid residues, said B being on the COOH-terminal side of any of said X^1, X^2, X^3 or X^4 amino acid residues, said B being on the NH_2-terminal side of any of said alanine or glycine residues, and said B being on the COOH-terminal side of any of said alanine or glycine residues.

28. The library of claim 23 wherein more than one said B is present.

29. The library of claim 28 wherein said Bs are adjacent to each other.

30. The library of claim 28 wherein said Bs are not adjacent to each other.

31. The library of claim 28 wherein said Bs are identical to each other.

32. The library of claim 28 wherein said Bs are not identical to each other.

33. The library of claim 1 wherein said displayed peptide sequence further comprises at least one constraint selected from the group consisting of a crosslink, a stacking interaction, a positive or negative charge, hydrophobicity, hydrophilicity, a structural motif and combinations thereof.

34. The library of claim 33 wherein said crosslink is a disulfide bond.

35. The library of claim 1 wherein said displayed peptide sequence further comprises at least one cysteine residue.

36. The library of claim 35 wherein said cysteine residue is selected from the group consisting of said cysteine residue being on the NH_2-terminal side of any of said X^1, X^2, X^3 or X^4 amino acid residues, said cysteine residue being on the COOH-terminal side of any of said X^1, X^2, X^3 or X^4 amino acid residues, said cysteine residue being on the NH_2-terminal side of any of said alanine or glycine residues, and said cysteine residue being on the COOH-terminal side of any of said alanine or glycine residues.

37. The library of claim 1 wherein at least one of said X^1, X^2, X^3 or X^4 residues is a cysteine residue.

38. The library of claim 1 further comprising at least one B, said B being a core binding sequence of p amino acid residues in length, and at least one cysteine residue.

39. The library of claim 38 wherein said displayed peptide sequence comprises:

$$(C/G)(Y^1)_{C^1}(C/G)(Y^2)_{C^2}B(C/G)(Y^3)_{C^3}(C/G)$$

wherein (C/G) is a cysteine or glycine residue.

40. The library of claim 1 wherein the complexity of said library is at least about 10^9.

41. An anchor library, comprising:

a collection of recombinant vectors,

each of said recombinant vectors having a nucleic acid sequence inserted in a gene, said nucleic acid sequence encoding a displayed peptide sequence,

said displayed peptide sequence of each of said vectors comprising

$$X^1(Y^1)_{c^1}X^2(Y^2)_{c^2}X^3(Y^3)_{c^3}X^4$$

wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein X^1 and X^4 are each attached to an amino acid residue that flanks said displayed peptide sequence, and

wherein said library does not contain more than about 10% of displayed peptide sequences different from said first mentioned displayed peptide sequences.

42. A method of making said anchor library of claim 1, comprising:

synthesizing a collection of nucleic acid sequences;

inserting said nucleic acid sequences into vectors to give recombinant vectors;

introducing said recombinant vectors into a host;

propagating said host having said recombinant vectors so as to result in a collection of recombinant vectors, each of said recombinant vectors having a nucleic acid sequence from said collection of nucleic acid sequences which encodes a displayed peptide sequence;

said displayed peptide sequence comprising:

$$X^1(Y^1)_{c^1}X^2(Y^2)_{c^2}X^3(Y^3)_{c^3}X^4$$

wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively c^1, c^2 and c^3 amino acid residues long and

any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein X^1 and X^4 are each attached to an amino acid residue that flanks said displayed peptide sequence, and

wherein at least about 10^5 to about 10^8 permutations of all possible permutations of said displayed peptide sequence are present in said anchor library.

43. The method of claim 42 wherein said library does not contain more than about 10% of displayed peptide sequences different from said first mentioned displayed peptide sequences.

44. A method of using said anchor library of claim 1 to identify a peptide sequence that binds to a target, comprising:

providing said anchor library of claim 1, said anchor library having a collection of recombinant vectors, each of said recombinant vectors having a nucleic acid sequence which encodes a displayed peptide sequence comprising

$$X^1(Y^1)_{c^1}X^2(Y^2)_{c^2}X^3(Y^3)_{c^3}X^4;$$

permitting the expression and display of said peptide sequence;

contacting said anchor library with said target under conditions in which said displayed peptide sequence binds to said target: and

identifying said displayed peptide sequence which binds to said target.

45. The method of claim 44 wherein said contacting step is by affinity purification.

46. The method of claim 44 wherein said identifying step comprises amplifying said recombinant vector which has said nucleic acid sequence which encodes for said displayed peptide sequence which binds to said target, and sequencing said nucleic acid sequence to determine said displayed peptide sequence which binds to said target.

47. The method of claim 44 further comprising the step of synthesizing said identified displayed peptide sequence.

48. The method of claim 44 wherein said target is selected from the group consisting of proteinaceous molecules and non-proteinaceous molecules.

49. The method of claim 44 wherein said target is selected from the group consisting of ligands, receptors, hormones, cytokines, antibodies, antigens, enzymes, enzyme substrates and viruses.

50. The method of claim 44 wherein said peptide sequence further comprises at least one B, said B being a core binding sequence of p amino acid residues in length.

51. The method of claim 44 wherein said peptide sequence further comprises at least one cysteine residue.

52. The method of claim 44 wherein at least one of said X^1, X^2, X^3 and X^4 residues is a cysteine residue.

53. A peptide identified by use of said library of claim 1 which peptide is useful as a diagnostic or therapeutic product in that said peptide is able to bind to a target molecule which is involved in a physiological process.

54. In a library having a collection of nucleic acid molecules encoding peptides having random amino acids, the improvement comprising a library in which the random amino acids are not continuous so that the amino acids in a peptide that are important contacts for interaction between said peptide and a target molecule can be identified.

55. In a library having a collection of nucleic acid molecules encoding peptides having random amino acids, the improvement comprising nucleic acid molecules encoding alanine or glycine or a combination of alanine and glycine residues in varying numbers acting as spacers between the random amino acids so that amino acid residues in a peptide that are important contacts for interaction between said peptide and a target molecule can be identified.

56. A collection of recombinant DNA molecules encoding peptide sequences having a plurality of different binding domains, said peptide sequences comprising:

$$X^1(Y^1)_{c^1}X^2(Y^2)_{c^2}X^3(Y^3)_{c^3}X^4$$

wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein X^1 and X^4 are each attached to an amino acid residue that flanks said peptide sequence, and

wherein at least about 10^5 to about 10^8 permutations of all possible permutations of said peptide sequence are present in said collection.

57. The collection of recombinant DNA molecules of claim 56 wherein said collection does not contain more than about 10% of displayed peptide sequences different from said first mentioned displayed peptide sequences.

58. The collection of recombinant DNA molecules of claim 56 wherein said peptide sequences are displayed on the surface of a biological material selected from the group consisting of a virus, phage, cell, spore and gene product.

59. A recombinant filamentous phage having a displayed peptide sequence with known binding properties,

said displayed peptide sequence being foreign to said filamentous phage,

said displayed peptide sequence comprising:

$$X^1(Y^1)_{c^1}X^2(Y^2)_{c^2}X^3(Y^3)_{c^3}X^4$$

wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein at least one of Y^1, Y^2 and Y^3 is at least about one amino

acid residue long, wherein X^1 and X^4 are each attached to an amino acid residue that flanks said displayed peptide sequence, and

said displayed peptide sequence being able to bind to a target.

60. A recombinant vector having a nucleic acid sequence inserted in a gene, said nucleic acid sequence encoding a displayed peptide sequence having known binding properties,

said displayed peptide sequence comprising:

$$X^1(Y^1)_{c^1}X^2(Y^2)_{c^2}X^3(Y^3)_{c^3}X^4$$

wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein at least one of Y^1, Y^2 and Y^3 is at least about one amino acid residue long, wherein X^1 and X^4 are each attached to an amino acid residue that flanks said displayed peptide sequence, and

said displayed peptide sequence being able to bind to a target.

61. A recombinant nucleic acid molecule having a nucleic acid sequence inserted in a gene, said nucleic acid sequence encoding a displayed peptide sequence having known binding properties,

said displayed peptide sequence comprising:

$$X^1(Y^1)_{c^1}X^2(Y^2)_{c^2}X^3(Y^3)_{c^3}X^4$$

wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein at least one of Y^1, Y^2 and Y^3 is at least about one amino acid residue long, wherein X^1 and X^4 are each attached to an amino acid residue that flanks said displayed peptide sequence, and

said displayed peptide sequence being able to bind to a target.

62. A recombinant protein having a displayed peptide sequence having known binding properties,

said displayed peptide sequence comprising:

$$X^1(Y^1)_{c^1}X^2(Y^2)_{c^2}X^3(Y^3)_{c^3}X^4$$

wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein at least one of Y^1, Y^2 and Y^3 is at least about one amino acid residue long, wherein X^1 and X^4 are each attached to an amino acid residue that flanks said displayed peptide sequence, and

said displayed peptide sequence being able to bind to a target.

63. An anchor library, comprising:

a collection of recombinant vectors,

each of said recombinant vectors having a nucleic acid sequence inserted in a gene, said nucleic acid sequence encoding a displayed peptide sequence,

said displayed peptide sequence of each of said vectors comprising

$$X^1(Y^1)_{c^1}X^2(Y^2)_{c^2}X^3(Y^3)_{c^3}X^4$$

wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is any specified amino acid or combination of specified amino acids that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein X^1 and X^4 are each attached to an amino acid residue that flanks said displayed peptide sequence, and

wherein at least about 10^5 to about 10^8 permutations of all possible permutations of said displayed peptide sequence are present in said anchor library.

64. The library of claim 63 wherein said library does not contain more than about 10% of displayed peptide sequences different from said first mentioned displayed peptide sequences.

65. The library of claim 63 wherein each Y^1, Y^2 and Y^3 is alanine or cysteine or a combination of alanine and cysteine.

66. The library of claim 63 wherein each Y^1, Y^2 and Y^3 is glycine or cysteine or a combination of glycine and cysteine.

67. An anchor library, comprising:

a collection of recombinant vectors,

each of said recombinant vectors having a nucleic acid sequence inserted in a gene, said nucleic acid sequence encoding a displayed peptide sequence,

said displayed peptide sequence of each of said vectors comprising

$$X^1(Y^1)_{C^1}X^2(Y^2)_{C^2}X^3(Y^3)_{C^3}X^4$$

wherein each X^1, X^2, X^3 and X^4 is an amino acid residue and any of X^1, X^2, X^3 and X^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a core binding sequence B of p amino acid residues in length or a combination of alanine and glycine or alanine and B or glycine and B, that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein X^1 and X^4 are each attached to an amino acid residue that flanks said displayed peptide sequence, and

wherein at least about 10^5 to about 10^8 permutations of all possible permutations of said displayed peptide sequence are present in said anchor library.

68. The library of claim 67 wherein said library does not contain more than about 10% of displayed peptide sequences different from said first mentioned displayed peptide sequences.

69. An anchor library, comprising:

a collection of recombinant vectors,

each of said recombinant vectors having a nucleic acid sequence inserted in a gene, said nucleic acid sequence encoding a displayed peptide sequence,

said displayed peptide sequence of each of said vectors comprising

$$Z^1(Y^1)_{c^1}Z^2(Y^2)_{c^2}Z^3(Y^3)_{c^3}Z^4$$

wherein each Z^1, Z^2, Z^3 and Z^4 is an amino acid residue or a core binding sequence B of p amino acid residues in length and any of Z^1, Z^2, Z^3 and Z^4 can be the same or different from any one other, wherein each Y^1, Y^2 and Y^3 is alanine or glycine or a combination of alanine and glycine that is respectively c^1, c^2 and c^3 amino acid residues long and any of Y^1, Y^2 and Y^3 if present can be the same or different from any one other, wherein each of c^1, c^2 and c^3 is 0 to about 20, wherein Z^1 and Z^4 are each attached to an amino acid residue that flanks said displayed peptide sequence, and

wherein at least about 10^5 to about 10^8 permutations of all possible permutations of said displayed peptide sequence are present in said anchor library.

70. The library of claim 69 wherein said library does not contain more than about 10% of displayed peptide sequences different from said first mentioned displayed peptide sequences.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/09383

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :436/518; 435/6, 320.1; 530/300, 350
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 436/518; 435/6, 320.1; 530/300, 350

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

None

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, CAS ONLINE
search terms: phage, bacteriophage, display, library, anchor, support, peptides

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	HART et al. Cell Binding and Internalization by Filamentous Phage Displaying a Cyclic Arg-Gly-Asp-containing Peptide. Journal of Biological Chemistry. 29 April 1994, Vol. 269, No. 17, pages 12468-12474, see entire document	1-70
Y	DEVLIN et al. Random Peptide Libraries: A Source of Specific Protein Binding Mloecules. Science. 27 July 1990, Vol. 249, pages 404-406, see entire document.	1-70

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 30 AUGUST 1996	Date of mailing of the international search report 18 SEP 1996
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer P. ACHUTAMURTHY Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/09383

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	FELICI et al. Mimicking of discontinuous epitopes by phage-displayed peptides, II. Selection of clones recognized by a protective monoclonal antibody against the Bordatella pertussis toxin from phage peptide libraries. Gene. 1993, Vol. 128, pages 21-27. see entire document.	1-70
Y	LENSTRA et al. Isolation of sequences from a random-sequence expression library that mimic viral epitopes. Journal of Immunological Methods. 1992, Vol. 152, pages 149-157. See entire document.	1-70

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/09383

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/09383

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

G01N 33/543; C07K 2/00, 14/00; C12Q 1/68; C12N 15/09, 15/70, 15/74

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s)1-52, and 63-70, drawn to anchor vector libraries, method of making and using said libraries, classified in class 436/518, and 435/6.
Group II, claim(s) 55, drawn to a peptide, classified in class 530/300+.
Group III, claim(s) 54-58, drawn to a collection of nucleic acid molecules, classified in class 435/6.
Group IV, claims 59-61, drawn to a phage/vector displaying a peptide sequence, classified in class 435/320.1.
Group V, claim 62, drawn to a recombinant protein, classified in class 530/350+.
The inventions listed as Groups I-V do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: The only linking feature among the claims is the peptide having the recited sequence. However, in the formula for this peptide sequence as shown, for example, in claim 1, the amino acids represented by 'Y' need not be present (because the values of c-1, c-2, or c-3 can be zero). This would give raise a collection of tetrapeptides without anyspecific requirements for the amino acids. Such peptides areobviously known in the prior art. Accordingly the linking technical feature is not "a special technical feature" as defined by PCT Rule 13.2, because it fails to make contribution over prior art.